ESTIMASI DATA HILANG
PADA RANCANGAN ACAK KELOMPOK LENGKAP

Tatik Widiharih
Program Studi Statistika, Jurusan Matematika FMIPA UNDIP
Jl. Prof. H. Soedarto, SH Tembalang Semarang 50275

Abstract. Randomized complete block design is a design to reduce the residual error in an experiment by removing variability due to a known and controllable nuisance variable. Missing observations introduce a new problem into the analysis since treatments are no longer orthogonal to blocks, that is, every treatment does not occur in every block, There are two general approaches to the missing values problem. The first is an exact analysis, the second is an approaximate analysis in which the missing observations are estimated and usual analysis of variance is performed just as if the estimated observations were real data, with the error degrees of freedom reduced by the number of missing observations. In this paper was discussed the second approach with completely analysis. Bigger's method is a simple method for estimating missing observations by using matrix approximation.

Key words: estimate, randomized complete block design, Biggers method, analysis of variance

1. PENDAHULUAN

Dalam suatu percobaan, kadang-kadang sulit didapatkan satuan percobaan yang relatif homogen. Pada masalah seperti ini satuan percobaan dapat dikelompokkan menurut satu arah, dua arah atau multi arah. Bila satuan percobaan tidak homogen dan pengelompokan dilakukan menurut satu arah maka digunakan ran-cangan acak kelompok lengkap ([2], [3], [4], [5], [6]).

Satuan percobaan yang digunakan kadang-kadang juga tidak dapat diukur responnya karena mati (untuk hewan/ tanaman, pasien meninggal), tabung reaksi pecah dan lain-lain. Kasus seperti ini kita berhadapan dengan kasus data hilang, dan data hilang ini akan menimbulkan masalah dalam analisis karena perlakuan dan kelompok menjadi tidak orthogonal. Untuk mengatasi masalah ini dapat dilakukan dengan dua cara yaitu :

1. Menganalisis data seadanya dengan menganggap rancangan yang digunakan adalah rancangan acak kelompok tidak lengkap.
2. Data yang hilang diestimasi dahulu, baru kemudian dilakukan analisis dengan konsekuensi derajat bebas galat (dengan sendirinya juga derajat bebas total) berkurang sejumlah data yang hilang.

Cara yang pertama biasanya tidak disukai karena lebih rumit, dan bagi pengguna statistika dengan dasar teori yang kurang akan menemui kesulitan [5]. Cara yang kedua biasa dilakukan oleh pengguna statistika. Metode yang biasa digunakan adalah metode Yates [5]. Metode ini dalam mengestimasi data hilang dengan prinsip meminimalkan jumlah kuadrat galat, Namun metode ini akan menjadi tidak menarik dan menjemukan jika data yang hilang banyak (lebih dari tiga) ([5], [6]).

Penyempurnakan metode Yates dengan pendekatan matriks dilakukan oleh [1] agar perhitungan menjadi lebih sederhana. Metode ini memang cukup tua, namun sampai saat ini tetap digunakan. Metode Yates maupun metode Biggers mendapatkan hasil estimasi yang sama. Metode Biggers hanya sebatas menentukan estimasi dari data yang hilang dan estimasi dengan metode ini akan menghasilkan bias untuk jumlah kuadrat perlakuan. Sehingga diperlukan penanganan khusus untuk menghilangkan bias tersebut.

Dalam tulisan ini dibahas dua hal yaitu cara mengestimasi data hilang dengan metode Biggers dan metode analisis yang diperlukan untuk menghilangkan bias. Analisis variansi untuk menangani bias ini dikenal dengan istilah analisis variansi alternatif. Setelah diperoleh tabel analisis variansi alternatif dilakukan uji lanjut alternatif dengan metode *Least Significance Difference* (LSD). Untuk memperjelas pembahasan diberikan contoh aplikasi untuk memperjelas pembahasan. Penghitungan dilakukan dengan paket SPLUS 2000 dan minitab 14.20.

2. RANCANGAN ACAK KELOMPOK LENGKAP (RAKL)

Suatu percobaan dengan menggunakan a buah perlakuan yang akan dicobakan dan masing-masing perlakuan dicobakan pada b kelompok yang berbeda, model liniernya adalah :

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}\,, \qquad (2.1)$$

$i=1,2,\ldots,a \quad j=1,2,\ldots,b$

dengan

Y_{ij} : pengamatan pada perlakuan ke i kelompok ke j,

μ : pengaruh rataan umum,

α_i : pengaruh perlakuan ke i,

β_j : pengaruh kelompok ke j,

ε_{ij} : komponen galat.

Bila diambil model tetap diasumsikan :

$$\sum_i \alpha_i = 0 \ldots\ldots \sum_j \beta_j = 0\,,$$

dan ε_{ij} berdistribusi normal dengan rata-rata nol dan variansi konstan $(\varepsilon_{ij} \sim N(0, \sigma^2)$.

Hipotesis yang dapat diambil adalah:

1. H_0 : $\alpha_i = 0$ untuk setiap i (tidak ada pengaruh perlakuan terhadap respon yang diamati).
   H_1 : paling sedikit ada satu i dengan $\alpha_i \neq 0$ (ada pengaruh perlakuan terhadap respon yang diamati).
2. H_0 : $\beta_j = 0$ untuk setiap j (tidak ada pengaruh kelompok terhadap respon yang diamati).
   H_1 : paling sedikit ada satu j dengan $\beta_j \neq 0$ (ada pengaruh kelompok terhadap respon yang diamati).

Rumus penghitungan untuk jumlah kuadrat total (*JK*T), jumlah kuadrat perlakuan (*JKP*), jumlah kuadrat kelompok (*JKK*) dan jumlah kuadrat galat (*JKG*) sebagai berikut.

$$JKT = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2 - \frac{Y_{..}^2}{a.b}\,,$$

$$JKP = \frac{1}{b}\sum_{i=1}^{a} Y_{i.}^2 - \frac{Y_{..}^2}{a.b}\,,$$

$$JKK = \frac{1}{a}\sum Y_{.j}^2 - \frac{Y_{..}^2}{a.b}\,.$$

$$JKG = JKT - JKP - JKK\,, \qquad (2.2)$$

dengan :

$$Y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}\,,$$

$$Y_{i.} = \sum_{j=1}^{b} Y_{ij}\,,$$

$$Y_{.j} \sum_{i=1}^{a} Y_{ij}.$$

2.1. Data Hilang dalam RAKL

Untuk memudahkan pembahasan digunakan notasi-notasi sebagai berikut.

X_{ij} : menyatakan data pada perlakuan ke i kelompok ke j hilang.

d_i : banyaknya data hilang pada perlakuan ke i.

c_j : banyaknya data hilang pada kelompok ke j.

Misalkan dalam penelitian ini ada p buah data hilang, maka :

$$p = \sum_{i=1}^{a} d_i = \sum_{j=1}^{b} c_j\,, \qquad (2.3)$$

dengan

T_i : total perlakuan ke i dengan d_i buah data hilang,

B_j : total kelompok ke j dengan c_j buah data hilang,

D : total seluruh pengamatan dengan p buah data hilang.

Data hilang pada kelompok yang sama dinamakan kelompok sekutu, dan jika data hilang pada perlakuan yang sama dinamakan perlakuan sekutu. Sehingga yang hilang data X_{gh}, akan mempunyai (c_h-1) kelompok sekutu, $(d_g\text{-}1)$ perlakuan sekutu dan $(p\text{-}(c_h+d_g\text{-}1)$ tanpa sekutu.

$$JKG = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2 - \frac{1}{b}\sum Y_{i.}^2 - \frac{1}{a}\sum_{j=1}^{b} Y_{.j}^2 + \frac{Y_{..}^2}{a.b}$$

$$= \sum_{i\neq(i)}\sum_{j\neq(J)} Y_{ij}^2 + \sum_{(i)}\sum_{(j)} X_{ij}^2$$

$$-\frac{1}{b}\left[\sum_i\left(\sum_j Y_{ij}\right)^2 + \sum_i\left(T_i + \sum_{(j)} X_{ij}\right)^2\right]$$

$$-\frac{1}{a}\left[\sum_j\left(\sum_i Y_{ij}\right)^2 + \sum_j\left(B_j + \sum_{(i)} X_{ij}\right)^2\right]$$

$$+\frac{1}{a.b}\left[D + \sum_{(i)}\sum_{(j)} X_{ij}\right]^2$$

$$= \sum_{(i)}\sum_{(j)} X_{ij}^2 - \frac{1}{b}\sum_i\left(T_i + \sum_{(j)} X_{ij}\right)^2$$

$$-\frac{1}{a}\sum_j\left(B_j + \sum_{(i)} X_{ij}\right)^2 \quad , \quad (2.4)$$

$$+\frac{1}{a.b}\left(D + \sum_{(i)}\sum_{(j)} X_{ij}\right)^2 + R$$

dengan R konstanta yang tidak mengan-dung X_{ij}.

2.2. Metoda Biggers

Pada dasarnya metode Biggers digunakan untuk mengestimasi untuk X_{ij} , yaitu $\hat{X}_{ij}$ ditentukan sedemikian sehingga JKG persamaan (2.4) minimum. Hal ini dilakukan dengan mengambil turunan dari JKG terhadap X_{ij} dan menyamakan dengan nol. Andaikan data yang hilang tersebut adalah X_{gh} ,

$$JKG = X_{gh}^2 - \frac{1}{b}\left(T_g + \sum_{(j)} X_{gj}\right)^2$$

$$-\frac{1}{b}\left(B_h + \sum_{(i)} X_{ih}\right)^2 + \frac{1}{a.b}\left(D + X_{gh}\right)^2 + R$$

$$\frac{\partial JKG}{\partial \hat{X}_{gh}} = 0$$

$$\Leftrightarrow a.b.\hat{X}_{gh} - a\sum_{(j)} X_{gj} - b\sum_{(i)} X_{ih} \qquad (2.5)$$

$$+\sum_{(i)}\sum_{(j)} X_{ij} = a.T_g + b.B_h - D$$

Persamaan (2.5) dikelompokkan dalam suku-suku yang berhubungan dengan kelompok sekutu, perlakuan sekutu dan tanpa sekutu sebagai berikut.

$$\Leftrightarrow a.b.\hat{X}_{gh} - a\left(\sum_{(j),i\neq g} X_{ij} + \hat{X}_{gh}\right)$$

$$-b\left(\sum_{(i),i\neq g} X_{ij} + \hat{X}_{gh}\right) + \left(\sum_{(i),i\neq g}\sum_{(j),j\neq h} X_{ij}\right.$$

$$\left. + \sum_{(i),i\neq g} X_{ih} + \sum_{(j),j\neq h} X_{gj} + \hat{X}_{gh}\right) = a.T_g$$

$$+b.B_h - D$$

$$\Leftrightarrow (a-1)(b-1)\hat{X}_{gh} + (1-a)\sum_{(j),j\neq h} X_{gj} + (1-b)\sum_{(i),i\neq g} X_{ih}$$

$$+\sum_{(i),i\neq g}\sum_{(j),j\neq h} X_{ij} = a.T_g + b.B_h - D$$

(2.6)

Analog untuk (p-1) data hilang yang lain. Sehingga diperoleh p buah persamaan yang analog dengan (2.5) dan (2.6). Bila ditulis dalam bentuk matriks

$A_{pxp}.\ X_{px1} = Q_{px1},$ (2.7)

dengan

- A_{pxp}: matriks simetri dengan elemen-elemen *(a-1)(b-1)* untuk kelompok dan perlakuan yang bersesuaian, *(1-a)* untuk perlakuan yang bersesuaian, *(1-b)* untuk kelompok yang bersesuaian dan 1 untuk lainnya. Matriks ini merupakan matriks nonsingular,
- X_{px1} : matriks dari data yang hilang,
- Q_{px1} : matriks nilai $a.T_g + b.B_h – D$ dari persamaan yang bersesuaian.

Dari persamaan (2.7) diperoleh :

$$\hat{X}_{px1} = A^{-1}.Q \qquad (2.8)$$

Untuk memperjelas matriks A_{pxp} , misalkan dalam percobaan ini ada 4 data yang hilang, yaitu : X_{kk} , X_{kl} , X_{mk} *dan* X_{st} Elemen-elemen dari A_{pxp} ditentukan sebagai berikut.

Subkrip	kk	kl	mk	st
kk	*(a-1)(b-1)*	*1-a*	*1-b*	*1*
kl	*1-a*	*(a-1)(b-1)*	*1*	*1*
mk	*1-b*	*1*	*(a-1)(b-1)*	*1*
st	*1*	*1*	*1*	*(a-1)(b-1)*

A X = Q

$$\begin{bmatrix} (a-1)(b-1) & 1-a & 1-b & 1 \\ 1-a & (a-1)(b-1) & 1 & 1 \\ 1-b & 1 & (a-1)(b-1) & 1 \\ 1 & 1 & 1 & (a-1)(b-1) \end{bmatrix} . \begin{bmatrix} X_{kk} \\ X_{kl} \\ X_{mk} \\ X_{st} \end{bmatrix} = \begin{bmatrix} a.T_k + b.B_k - D \\ a.T_k + b.B_l - D \\ a.T_m + b.B_k - D \\ a.T_s + b.B_t - D \end{bmatrix}$$

2.3. Analisis Variansi Alternatif.

Untuk mengatasi bias dilakukan analisis variansi alternatif ([6]), dengan langkah-langkah sebagai berikut.

1. Dari data seadanya (data tidak lengkap karena beberapa data hilang), dihitung

$$JKT^* = \sum_{i,j} Y_{ij}^2 - \frac{\left(\sum_{i,j} Y_{ij}\right)^2}{N},$$

$$JKK^* = \sum_j \frac{Y_{.j}^2}{n_j} - \frac{\left(\sum_{i,j} Y_{ij}\right)^2}{N}, \qquad (2.9)$$

$$N = \sum_j n_j.$$

dimana n_j adalah banyaknya perlakuan yang muncul (dicobakan) pada kelompok ke-j. *JKK** ini selanjutnya disebut sebagai jumlah kuadrat kelompok yang mengabaikan perlakuan.

2. Setelah data hilang diestimasi, dimasukan data tersebut bersesuaian dengan data hilang kemudian dihitung : jumlah kuadrat galat (JKG) menggunakan persamaan (2.2).
3. Selanjutnya dihitung jumlah kuadrat perlakuan setelah dikoreksi terhadap kelompok (JKP*) dengan rumus : *JKP* = JKT* - JKK* - JKG*. Akan diperoleh tabel anova alternatif seperti pada Tabel 1.

$$JKT^* = \sum_{i,j} Y_{ij}^2 - \frac{\left(\sum_{i,j} Y_{ij}\right)^2}{N},$$

$$JKK^* = \sum_j \frac{Y_{.j}^2}{n_j} - \frac{\left(\sum_{i,j} Y_{ij}\right)^2}{N}, \qquad (2.10)$$

$$N = \sum_j n_j.$$

dimana n_j adalah banyaknya perlakuan yang muncul (dicobakan) pada kelompok ke *j*. *JKK** ini selanjutnya disebut sebagai jumlah kuadrat kelompok yang mengabaikan perlakuan.

Apabila ada pengaruh perlakuan dilakukan uji lanjut dengan galat baku selisih perlakuan ke *i* dan ke *j* :

$$S_{\overline{Y}} = \sqrt{KTG\left(\frac{1}{n_i^*} + \frac{1}{n_j^*}\right)},$$

dengan n_i^* dan n_j^* disebut dengan ulangan efektif dengan aturan :

$$n_i^* = \begin{cases} 1, \text{untuk kelompok dengan perlakuan ke i dan} \\ \quad \text{ke j ada} \\ \dfrac{a-2}{a-1}, \text{untuk kelompok dengan perlakuan ke.i ada} \\ \qquad \text{perlakuan ke j hilang} \\ 0, \text{untuk kelompok dengan perlakuan ke i hilang} \end{cases}$$

$$LSD = t_{\frac{\alpha}{2};ab-a-b+1-p} \cdot S_{\overline{Y}}$$

Rata-rata perlakuan ke i dan ke j dikatakan berbeda jika $\left|\overline{Y}_{i.} - \overline{Y}_{j.}\right| > LSD$ (rata-rata perlakuan berdasarkan data lengkap (data hilang yang diestimasi telah dimasukan))

3. STUDI KASUS

Suatu percobaan dengan menggunakan 6 perlakuan berbeda yang dicobakan pada 4 kelompok berbeda, dimana 4 data diantaranya hilang seperti pada Tabel 2.

$$\hat{X}_{4x1} = A^{-1}.Q$$

$$\begin{bmatrix} \hat{X}_{13} \\ \hat{X}_{21} \\ \hat{X}_{22} \\ \hat{X}_{34} \end{bmatrix} = \begin{bmatrix} 15 & 1 & 1 & 1 \\ 1 & 15 & -5 & 1 \\ 1 & -5 & 15 & 1 \\ 1 & 1 & 1 & 15 \end{bmatrix}^{-1} . \begin{bmatrix} 88 \\ 66{,}8 \\ 64{,}8 \\ 84{,}6 \end{bmatrix}$$

Dengan menggunakan paket SPLUS 2000 diperoleh :

$$\begin{bmatrix} \hat{X}_{13} \\ \hat{X}_{21} \\ \hat{X}_{22} \\ \hat{X}_{34} \end{bmatrix} = \begin{bmatrix} 4{,}80980 \\ 5{,}692308 \\ 5{,}592308 \\ 4{,}567033 \end{bmatrix}$$

Dengan menggunakan paket program minitab 14.20 diperoleh :
JKT^* = 27,1920 JKK^* = 1,18
JKG = 7,9661 dan JKP^* = 18,0459
Diperoleh table anova alternative seperti Tabel 3.

Berdasarkan Table 3, dapat disimpulkan ada pengaruh perlakuan terhadap respon yang diamati, selanjutnya dilakukan uji lanjut dengan $\alpha = 5\%$ seperti table 4.

Normalitas dari residual juga terpenuhi, dari output minitab 14.20 diperoleh statistic hitung 0,093 dengan p-value > 0,15. Demikian juga homogenitas residual statistic hitung 6,34 dengan p-value 0,275.

3. KESIMPULAN

Dalam kasus penelitian dengan beberapa data hilang diperlukan penanganan khusus, tentunya dengan analisis yang lebih rumit. Dalam estimasi data hilang sebaiknya menggunakan bantuan paket program computer untuk memperoleh hasil yang lebih teliti dan memudahkan perhitungan. Analisis yang digunakan untuk mengatasi bias adalah analisis variansi alternatif. Uji lanjut LSD juga mempunyai bentuk yang khusus yang lebih rumit. Bagi peneliti dituntut berhati-hati sehingga semua data bisa teramati dan analisis menjadi sederhana.

4. DAFTAR PUSTAKA.

[1]. Biggers, J.D. (1959). *The Estimation of Missing and Mixed-up Observations in Several Experimental Designs.* Biometrics, 91–105.

[2]. Gasperzs, V. (1992). *Teknik Analisis Dalam Penelitian Percobaan 2.* Tarsito, Bandung.

[3]. Gomez, K.A & Gomez, A.A. (1984), *Statistical Procedures for Agricultural Research 2nd* ,John Willey & Sons Inc. Singapore.

[4]. Haslet, *et al* . (1997). *Experimental Design for Researchers,* Departement of Statistics, Faculty of Information and Mathematical Science, Massey University.

[5]. Montgomery, D.C. (2006). *Design and Analysis of Experiment 6nd*, John Willey & Sons Inc. New York.

[6]. Steel, R.G.D and Torrie, J.H. (1989). *Prinsip dan Prosedur Statistika : Suatu Pendekatan Biometrik Edisi 1,* PT Gramedia Pustaka Utama Jakarta. (Alih Bahasa : Ir. Bambang Sumantri)

DAFTAR TABEL

Tabel 1. Tabel Anova Alternatif RAKL dengan p Buah Data Hilang.

Sumber Keragaman	Derajat bebas	Jumlah Kuadrat	Kuadrat tengah	F_{hitung}	F_{tabel}
1. Kelompok mengabaikan perlakuan	*b-1*	*JKK**	*KTK*=JKK*/(b-1)*	*KTK*/KTG*	$F_{(b-1);(ab-a-b+1-p)}(\alpha)$
2. Perlakuan terkoreksi	*a-1*	*JKP**	*KTP*=JKP*/(a-1)*	*KTP*/KTG*	$F_{(a-1);(ab-a-b+1-p)}(\alpha)$
3. Galat	*ab-a-b+1-p*	*JKG*	*KTG=JKG/(ab-a-b+1-p)*		
Total	*ab-1-p*	*JKT**			

Kriteria uji : tolak H_0 jika $F_{hitung} > F_{tabel}$ yang bersesuaian.

Tabel 2. Hasil Pengamatan RAK dengan 6 Perlakuan, 4 Kelompok dimana 4 Data hilang

Perlakuan	kelompok				T_i
	1	2	3	4	
1	4,4	5,9	…..	4,1	14,4
2	…..	…..	4,9	7,1	12,0
3	4,4	4,0	4,5	……	12,9
4	6,8	6,6	7,0	6,4	26,8
5	6,3	4,9	5,9	7,1	24,2
6	6,4	7,3	7,7	6,7	28,1
B_j	28,3	28,7	30,0	31,4	D=118,4

Sumber data [6]

Tabel 3. Tabel Anova Alternatif RAK dengan dengan 6 Perlakuan, 4 Kelompok dimana 4 Data hilang

Sumber Keragaman	Derajat bebas	Jumlah Kuadrat	Kuadrat tengah	F_{hitung}	F_{tabel}
1. Kelompok mengabaikan perlakuan	3	1,18	0,3933	0,5431	3,59
2. Perlakuan terkoreksi	5	18,0459	3,6092	4,9837	3,20
3. Galat	11	7,9661	0,7242		
Total	19	27,1920			

Tabel 4. Uji Lanjut LSD Untuk Semua Pasangan Rata-rata.

no	perlakuan i dan j	ni*	nj*	$S_{\bar{Y}}$	LSD	$\left\|\bar{Y}_{i.} - \bar{Y}_{j.}\right\|$	kesimpulan
1	1 dan 2	2.6	1.8	0.82515	2.885548	1.01868	tidak berbeda
2	1 dan 3	2.6	2.6	0.746376	2.610076	0.43571	tidak berbeda
3	1 dan 4	2.6	3.8	0.684921	2.39517	1.89753	tidak berbeda
4	1 dan 5	2.6	3.8	0.684921	2.39517	1.24753	tidak berbeda
5	1 dan 6	3	3.8	0.657251	2.298407	2.2225	tidak berbeda
6	2 dan 3	1.8	2.6	0.82515	2.885548	1.45439	tidak berbeda
7	2 dan 4	2	3.6	0.750511	2.624537	0.87885	tidak berbeda
8	2 dan 5	2	3.6	0.750511	2.624537	0.22885	tidak berbeda
9	2 dan 6	2	3.6	0.750511	2.624537	1.20385	tidak berbeda
10	3 dan 4	3	3.8	0.657251	2.298407	2.3332	berbeda
11	3 dan 5	3	3.8	0.657251	2.298407	1.68324	tidak berbeda
12	3 dan 6	3	3.8	0.657251	2.298407	2.65824	berbeda
13	4 dan 5	4	4	0.601747	2.104311	0.65	tidak berbeda
14	4 dan 6	4	4	0.601747	2.104311	0.325	tidak berbeda
15	5 dan 6	4	4	0.601747	2.104311	0.975	tidak berbeda